Coordinate-Free Linear Algebra

Nic Ford

1 Introduction

All of the content in this article is about linear algebra, but it's not an article about learning linear algebra. Rather, we are going to use linear algebra as a vehicle for introducing a perspective that pervades a good chunk of modern mathematics, but which isn't encountered much even by people working in other technical fields.

There are many constructions in linear algebra — the determinant is a great example to think about for this — which are usually built in two steps. First, you pick a basis for whatever vector space you're working with and define your chosen object in terms of that basis. Then, with this definition in hand, you prove that it actually didn't matter which basis you picked; you get the same thing out either way.

This is a bit backwards. If the choice didn't actually affect the final result of the computation, then why did we have to make a choice at all? Take the determinant as an example. When it's defined in this way it's usually given as a giant sum, each term of which involves products of various entries of the matrix. If you prove that the determinant doesn't depend on the choice of basis, you will see that changing the basis can and does change each individual term of this sum. It is *only* the final result that stays the same.

It's possible (and indeed common) to present the determinant in a way that makes all of this look like a giant coincidence. When something like this happens, we should take it as a sign that the definition we started with isn't capturing the concept at its most fundamental level; the giant sum may give you a formula for the determinant, but it doesn't tell you much about the geometric idea that the determinant is meant to capture, about what the determinant "really is."

This article is all about what do to about this. We plan to rebuild a big chunk of linear algebra from the ground up while remaining **coordinate-free**, that is, never defining anything in terms of a choice of basis. Often when we already have a definition in hand we will be able to go back and see what it looks like in a particular basis — it is useful, for example, to know that the trace of a matrix is the sum of its diagonal entries — but the definition itself will never depend on such a choice. When I first learned to look at these constructions in this way I found that it did a lot to help me understand where they come from and why they're important, and I hope that when you're doing reading you will feel the same way.

Many of the tools we will use to do this come from a part of mathematics called "category theory." It won't be necessary to know a single thing about category theory to follow this article. (In fact, I won't even be telling you what a category is!) But if your interests lie in that direction the examples in this article may help explain the intuition behind some of its basic definitions.

This article grew out of a class called "Multilinear Algebra" that I taught at Canada/USA Mathcamp in 2016. My notes for that class are available with the rest of my Mathcamp notes,

but there shouldn't be anything there that I'm not talking about here. I'm indebted to Asilata Bapat for the idea for that class, and therefore ultimately for this article; she taught a Mathcamp class with the same title in 2013 which was the inspiration for my version. I am also thankful to Jake Levinson for several fruitful conversations about the organization and content of this article.

The target audience for this article is someone who is familiar with linear algebra, and in particular with the definition and basic properties of vector spaces. It will help if you've seen the usual, basis-dependent presentation of most of the ideas we'll be talking about, like direct sums, transposes, traces, and determinants. (The exception to this is the tensor product, which I will be introducing without assuming that it's at all familiar.)

I will be referring to some things using terminology that might be more familiar to mathematicians that to people who use linear algebra in other technical fields. Specifically:

- I will often just say "map" to mean what is sometimes called a "linear map" or "linear transformation." Later in the article we will discuss something called a "bilinear map." Bilinear maps are not linear maps, and so I'll always refer to them by that full name, never just "maps."

- I will often use notation like $f : V \to W$. This means that f is a linear map from V to W. This is never meant to imply that f is necessarily surjective, that is, the image of f is not necessarily all of W.

- I will write composition of linear maps using a notation that looks like multiplication, so fg is the map that takes v to $f(g(v))$. (If you've picked bases and written f and g as matrices, this corresponds to multiplying the matrices.)

- I will use the word "kernel" to refer to the set of vectors that a linear map sends to 0. (Some fields call this is called the "null space" of the map.)

- I will use "vector" interchangeably with "element of a vector space."

- I will often refer to invertible linear maps as "isomorphisms."

Finally, we will be using the concept and some basic properties of the quotient of a vector space by a subspace. If you're not comfortable with this already there is an exercise in the first section to catch you up.

2 Universal Properties

In this section we'll introduce the concept of a universal property with a simple example: the direct sum of two vector spaces. The direct sum is simpler than the examples we'll spend most of our time with, and so it provides a good setting for learning how to work with universal properties and what they're good for.

Recall that, given two vector spaces V and W, we can form their **direct sum** $V \oplus W$ by taking the set of ordered pairs $\{(v, w) : v \in V, \ w \in W\}$ and defining addition and scalar multiplication componentwise, that is,
$$(v, w) + (v', w') = (v + v', w + w')$$
and
$$\lambda(v, w) = (\lambda v, \lambda w).$$

There is a way to think about direct sums that gets at the reason the construction is important: $V \oplus W$ consists of every vector from V, every vector from W, and their sums. In other words, it's a vector space and it has a copy of V and a copy of W, but we don't have to insist on anything else — the fact that it must then contain sums of vectors from V and W just follows from the fact that it's a vector space.
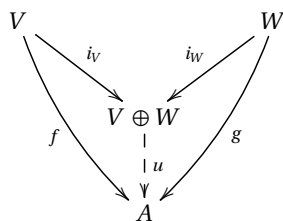
Our goal in this section will be to find a way to characterize the direct sum that more directly captures this intuition. First, we can express the idea that $V \oplus W$ "contains a copy of V and W" through the existence of two linear maps, which we'll call $i_V : V \to V \oplus W$ and $i_W : W \to V \oplus W$. (They are given by $i_V(v) = (v, 0)$ and $i_W(w) = (0, w)$. The i stands for *inclusion*.) We'll draw these two maps in a diagram like this:

$$
\begin{array}{ccc}
V & & W \\
 & \searrow^{i_V} \quad {}^{i_W}\swarrow & \\
 & V \oplus W &
\end{array}
$$
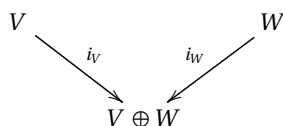
These maps capture the first essential fact about $V \oplus W$: that it contains vectors from V and vectors from W. But the existence of i_V and i_W doesn't uniquely specify $V \oplus W$, and so this isn't enough to serve as a definition. Indeed, you can build a map from V and a map from W to *any* vector space, including the zero space. Insisting that the maps be injective isn't enough either: that would be satisfied by, say, $V \oplus W \oplus V$ with $i_V(v) = (v, 0, 0)$ and $i_W(w) = (0, w, 0)$. For that matter, given any linear map $f : V \to W$, injective or not, we could stick W in the position of the direct sum by taking i_V to just be f and i_W to be the identity. This isn't what we want either, because it introduces extra linear equations that are satisfied by the images of i_V and i_W, something that doesn't happen for the real direct sum.

So what we're after isn't just a vector space we can map V and W into. Among all vector spaces with maps like i_V and i_W, the direct sum is special for two reasons: it doesn't contain any extra elements beyond those forced by the definition of a vector space, and it doesn't have any unnecessary linear relations among the elements it contains. What we're aiming for is a way to say not just that $V \oplus W$ has these maps from V and W, but that it's the "most general" vector space with these maps. It's this idea that's meant to be captured by the following fact:

Suppose there is some vector space A and linear maps $f : V \to A$ and $g : W \to A$. Then there is a unique linear map $u : V \oplus W \to A$ so that $f = u i_V$ and $g = u i_W$. This property of $V \oplus W$ is summarized by the following diagram:

$$
\begin{array}{ccc}
V & & W \\
 & \searrow^{i_V} \quad {}^{i_W}\swarrow & \\
 & V \oplus W & \\
 {}^{f}\searrow & \Big\downarrow {\scriptstyle u} & {}^{g}\swarrow \\
 & A &
\end{array}
$$

Whenever we draw a diagram like the one in this proposition, we'll usually want to say that any two ways of getting from one vector space in the diagram to another are equal as linear maps. In this example, there are two ways of getting from V to A: you can follow f directly, or you can apply i_V and then u. The conclusion of the theorem is, in part, that there is a u that makes these two paths the same, that is, makes $f = u i_V$. Whenever this happens for every pair of vector spaces in the diagram (as it does here) we'll say that the diagram **commutes**.

One helpful way to look at the statement of the proposition is to see that it gives a one-to-one correspondence between linear maps out of $V \oplus W$ and *pairs* of linear maps out of V and W. The existence of i_V and i_W alone gives you a way to turn a map $u : V \oplus W \to A$ into a pair of maps $f : V \to A$ and $g : W \to A$ — just set $f = ui_V$ and $g = ui_W$ — and the universal property says that this process can be inverted, giving a way to produce u given the existence of f and g.

The solid and dashed lines in the diagram are meant to encode which maps are assumed to exist ahead of time and which one the proposition produces: the proposition says that, whenever the solid lines exist, we may conclude that the dashed line also exists (and that it makes the diagram commute).

Let's prove that the direct sum satisfies this property. Indeed, the definition of u is forced by the conditions we're imposing on it: if $ui_V = f$, then we know that any vector of the form $(v, 0)$ has to map to $f(v)$, since we need $f(v) = u(i_V(v)) = u((v, 0))$. Similarly, any $(0, w)$ has to map to $g(w)$. But then, since u has to be linear, this forces

$$u((v, w)) = u((v, 0) + (0, w)) = u((v, 0)) + u((0, w)) = f(v) + g(w).$$

And it's straightforward to check that this choice of u is indeed linear. Note that this gives us both the existence and the uniqueness: we have built a map satisfying the rule and we've also shown that it's the only possible such map.

This completes the proof, but we should also talk about why the universal property matches the intuition we were aiming for. I mentioned two ways in which a vector space with maps like i_V and i_W might fail to be the direct sum: it might have extra elements, or there might be extra linear equations satisfied by the elements. What we'd like to show is that if we have some vector space E with maps $i_V : V \to E$ and $i_W : W \to E$ that either has extra elements or satisfies extra linear equations, then it can't be put in the position of $V \oplus W$ (so we say that E "doesn't satisfy the universal property").

The property expressed in our diagram is indeed enough to exclude both of these possibilities. We'll take them one at a time, starting with the linear equations. Suppose we have some vector space E with maps $i_V : V \to E$ and $i_W : W \to E$ that fit into that diagram; the claim is that if we have $i_V(v) = i_W(w)$ for some v and w, then E doesn't satisfy the universal property (except if v and w are both zero). To see this, it's enough to find *any* vector space A with maps $f : V \to A$ and $g : W \to A$ where $f(v) \neq g(w)$, because then it is impossible to find a map $u : E \to A$ with the properties we want. (I encourage you to convince yourself of this right now if it's not clear.) And indeed, if v and w are not both zero, there will always be such an A; the direct sum itself gives an example! Note that this also forces i_V and i_W to be injective: if, say, $i_V(v) = 0$ and $v \neq 0$, we can take $w = 0$ in the above argument.

If instead there were some $x \in E$ which was not of the form $i_V(v) + i_W(w)$, we would have a different problem: given a vector space A with maps $f : V \to A$ and $g : W \to A$, the fact that $f = ui_V$ and $g = ui_W$ would do nothing to constrain the value of $u(x)$, and indeed there would be a map $u : E \to A$ sending x anywhere we want. So in this case we would violate the universal property not because u doesn't exist but because u isn't unique.

So our universal property gives us a precise way to characterize the direct sum the way we set out to, as the most general vector space with a map from V and from W. Implicit in this claim, though, is something that we haven't checked yet: that there's only one vector space that satisfies the universal property.

We should be careful about what we mean by "only one." We can't literally mean that $V \oplus W$ is the only vector space that satisfies the universal property; $W \oplus V$, for example, satisfies it just as well. Rather, we mean that any vector space that does so is *isomorphic* to $V \oplus W$, that is,

if a vector space F satisfies the universal property, then there is an invertible linear map from $V \oplus W$ to F, giving us a way to identify all the vectors in F with corresponding vectors in $V \oplus W$.

The argument for uniqueness is very important to our larger discussion of universal properties, so we'll go through it in detail. Suppose there are two vector spaces, E and F, that both satisfy the universal property. Call the corresponding inclusion maps $i_V : V \to E$, $i_W : W \to E$, $j_V : V \to F$, and $j_W : W \to F$. We can use these maps and the fact that E satisfies the universal property to get a map $u : E \to F$ making this diagram commute (so $ui_V = j_V$ and $ui_W = j_W$):
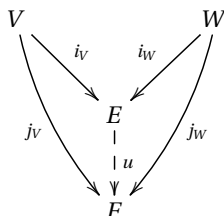
But F *also* satisfies the universal property, so we can switch the roles of E and F and get a map u' pointing the other way (so $u'j_V = i_V$ and $u'j_W = i_W$):
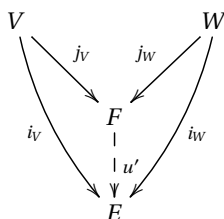
Now, consider the map $u'u : E \to E$. First, note that the commuting diagrams tell us that $u'ui_V = u'j_V = i_V$, and similarly $u'ui_W = i_W$. But this lets us set up yet another diagram using $u'u$, this time with E in both middle positions:
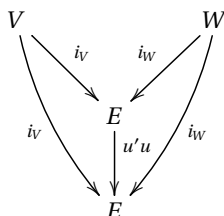
We use this to invoke the universal property one last time: if we had placed the identity map on E in place of $u'u$, we would also have made the diagram commute, so by the uniqueness portion of the universal property we must have that $u'u$ is the identity! An exactly similar argument shows that uu' is the identity map on F. These two facts together mean that $u' = u^{-1}$ — we have our invertible linear map, which completes the proof.

There are several features of this argument that should be highlighted.

1. Notice that we never used anything about the actual structure of the direct sum in the proof of uniqueness. The universal property all by itself is enough to guarantee that there can be at most one vector space that satisfies it. This is a general feature of universal properties, and it's part of what makes this perspective so powerful.

2. The "at most" in "at most one" is important though: defining a universal property doesn't automatically produce an object that satisfies it. We will discuss this more in the next section: the tensor product, unlike the direct sum, is an object whose universal property is much easier to describe than the construction itself.

3. Placing $V \oplus W$ in the role of E, the proof shows that, given any F satisfying the universal property, the invertible map $u : V \oplus W \to F$ is *itself* unique. So not only is the direct sum the only vector space that fits in the diagram, but there is only one way to "line up" the elements of $V \oplus W$ with the elements of F, namely $u((v, w)) = j_V(v) + j_W(w)$.

4. Since u is unique, the fact that $u i_V = j_V$ and $u i_W = j_W$ tells us that if F is going to satisfy the universal property there is also only one choice for the maps j_V and j_W. (If $E = V \oplus W$ as above, I encourage you to check that this unique choice works out to $j_V(v) = u((v, 0))$ and $j_W(w) = u((0, w))$.) So it is not only the vector space that is unique but also these two maps.

All of these facts — that the universal property specifies an object uniquely up to a unique invertible map, and that the maps attached to the object are unique as well — are going to be true of all of the universal constructions we talk about in this article, and for basically the same reasons.

None of these facts, of course, is especially hard to prove directly from the more concrete definition of the direct sum we started with. Instead, I encourage you to think of the universal property perspective as practice for the other, more complicated universal properties we discuss later on. By seeing how this tool works in this simple setting, where there's not as much going on under the hood, we will be better prepared for a situation where the proofs with diagrams really are simpler than the proofs with symbols.

Exercises

The easiest (and maybe only) way to get comfortable with these sorts of universal constructions is to get some practice working with them yourself. I've therefore decided to end most of the sections of this article with exercises. Feel free to send me an e-mail with any questions you might have.

1. Suppose we have two linear maps, $f : V \to V'$ and $g : W \to W'$. We'll write $f \oplus g$ for the map from $V \oplus W$ to $V' \oplus W'$ defined by setting

$$(f \oplus g)((v, w)) = (f(v), g(w)).$$

(a) Prove that the following diagram commutes:

Given this, how could we have deduced the existence of $f \oplus g$ and the above formula purely from the universal property of the direct sum?

(b) Prove that taking the direct sum of maps in this way respects composition. That is, given linear maps $f : V \to V'$, $g : W \to W'$, $f' : V' \to V''$, and $g' : W' \to W''$, we have

$$(f' \oplus g')(f \oplus g) = f'f \oplus g'g.$$

Try to prove this in two ways: once directly from the definition of $f \oplus g$ in terms of ordered pairs, and once only using the universal property.

(c) Suppose V and W are finite-dimensional, with bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$. We can form a basis for $V \oplus W$ by taking the $i_V(v_i)$'s in order followed by the $i_W(w_i)$'s, and similarly for $V' \oplus W'$. If we do this, how does the matrix for $f \oplus g$ relate to the matrices for f and g?

2. Suppose we're given a linear map $f : V \to W$. Consider the following universal property: we ask for a vector space K and a linear map $e : K \to V$ so that $fe = 0$, and so that for any other linear map $m : M \to V$ with $fm = 0$, there is a unique map $u : M \to K$ making the following diagram commute:

(When we say that this diagram commutes, we aren't saying that $f = 0$, but all other chains of arrows should be equal, including that $eu = m$. Note that any map composed with 0 is itself 0, so, for example, the first row commuting is what gives us that $fe = 0$.)

(a) Prove that this universal property uniquely specifies K up to a unique isomorphism, just like we did for the direct sum.

(b) I claim that this universal property is satisfied by an object you're already familiar with. What is it?

3. In this problem, we'll look at the universal property corresponding to the diagram from the previous problem with all the arrows reversed (and some of the vector spaces renamed):

We're looking for a vector space Q with a map $p : W \to Q$ so that $pf = 0$ and, whenever there's some $m : W \to M$ with $mf = 0$, there's a unique $u : Q \to M$ with $up = m$. We'll call Q (together with the associated map p) the **cokernel of** f.

(a) You can skip this part if you're already familiar with the concept of quotient vector spaces.

Suppose you're given a vector space A and a subspace $B \subseteq A$. We're going to construct a new vector space, called the **quotient space**, written A/B. The idea is that A/B will look like A, except that every element of B has been identified with 0.

We put an equivalence relation on A as follows: say that $a \sim a'$ if there's some $b \in B$ with $a' = a + b$. Then, as a set, A/B is the corresponding set of equivalence classes.

Prove that there is a consistent way to define addition and scalar multiplication on A/B, making it into a vector space. Also, show that the function $p : A \to A/B$ taking an element of A to its equivalence class is a surjective linear map and that $\ker p = B$.

(b) Suppose B is a subspace of A and $p : A \to A/B$ is the map taking each element of A to is equivalence class. Show that for any linear map $f : A \to C$ for which $f(B) = 0$, there is a unique linear map $g : A/B \to C$ so that $gp = f$. In this case, we'll say that f **descends to** A/B**.**

(c) Use the preceding exercises to construct a Q and p satisfying the universal property.

(d) Suppose V and W are finite-dimensional. What is the dimension of Q?

3 The Tensor Product

We are now ready to talk about the tensor product, the construction that we'll be using in one way or another for the rest of this article. Unlike with the examples from the last section, I'm not going to assume that you're already familiar with what a tensor product is or how it works. Even if you are, though, you'll hopefully still learn something by looking at them from this new perspective. The tensor product underlies many foundational ideas in linear algebra; we'll focus in this article on the trace and the determinant, but these are far from the only examples.

3.1 Bilinear Maps and the Tensor Product

In a similar way to how we defined the direct sum of V and W in terms of pairs of maps out of V and W, the tensor product is defined in terms of what are called "bilinear maps." We'll start by recalling the definition.

Let V, W, and A be vector spaces. A **bilinear map from** V **and** W **to** A is a function q from $V \times W$ to A satisfying the following properties:

- For any $v, v' \in V$ and $w \in W$,

$$q(v + v', w) = q(v, w) + q(v', w).$$

- For any $v \in V$ and $w, w' \in W$,

$$q(v, w + w') = q(v, w) + q(v, w').$$

- For any $v \in V$, $w \in W$, and $\lambda \in \mathbb{R}$,

$$q(\lambda v, w) = q(v, \lambda w) = \lambda q(v, w).$$

When $V = W$ and $A = \mathbb{R}$, so that q goes from $V \times V$ to $\mathbb{R}$, we call q a **bilinear form on** V. Any inner product on V is a bilinear form, in particular the usual dot product on $\mathbb{R}^n$.

Even though we can make $V \times W$ into a vector space — by treating it as a direct sum — it's important to note that this doesn't make q a linear map! We can see this even by looking at the dot product on $\mathbb{R}^2$: let $v = (1, 1)$ and $w = (1, 0)$. Then q sends (v, w) to 1, but it sends $2(v, w) = ((2, 2), (2, 0))$ to 4. Instead, the intuition to have is that a bilinear map is a map that is "linear in each coordinate separately": if I pick a single $v \in V$, then the map $q_v : W \to A$ given by $q_v(w) = q((v, w))$ is linear, and similarly the other way.

The **tensor product** is a vector space that has the same relationship to bilinear maps out of V and W that the direct sum has to pairs of linear maps out of V and W. That is, it's a vector space $V \otimes W$ and a bilinear map $p : V \times W \to V \otimes W$ that together satisfy a universal property: given any other bilinear map $q : V \times W \to A$, there is a unique *linear* map $u : V \otimes W \to A$ so that $q = up$, that is, u makes the following diagram commute:

$$
\begin{array}{ccc}
V \times W & \xrightarrow{\ p\ } & V \otimes W \\
 & \searrow{\scriptstyle q} & \Big\downarrow{\scriptstyle u} \\
 & & A
\end{array}
$$

(In this article we will always indicate bilinear maps in diagrams with double-bodied arrows like this. But be aware that I made this notation up; it is not at all standard!) Just as we could interpret the universal property of the direct sum as giving a one-to-one correspondence between maps out of $V \oplus W$ and pairs of maps out of V and W, we can interpret this one as giving a way to turn a *bilinear* map out of V and W into a *linear* map out of $V \otimes W$, and vice versa.

It is important to emphasize, though, that we don't automatically know that there actually is an object that satisfies this universal property. When discussing the universal property of the direct sum we already had a vector space in mind that would end up satisfying it, but in this case we have yet to actually construct one.

It will turn out that we can do this — every pair of vector spaces does in fact have a tensor product — but this will wait until the next subsection. In the exercises you will once again formulate and prove a uniqueness statement for this universal property: given any two vector spaces that fit in the position of $V \otimes W$ in this universal property, there is a unique isomorphism between them making the appropriate diagram commute.

This definition is pretty abstract, so let's look at what it tells us about the elements of $V \otimes W$. The existence of the bilinear map p means that for every $v \in V$ and $w \in W$, there's a corresponding element of $p((v, w)) \in V \otimes W$. We'll call this element $v \otimes w$. The fact that p is a bilinear map gives us some relations among these elements:

$$(v + v') \otimes w = v \otimes w + v' \otimes w, \quad v \otimes (w + w') = v \otimes w + v \otimes w',$$

and for any scalar λ,

$$(\lambda v) \otimes w = v \otimes (\lambda w) = \lambda(v \otimes w).$$

(The fact that $\otimes$ distributes over addition is why we use a symbol that's so suggestive of multiplication.) In particular, taking $\lambda = 0$, we get that $0 \otimes w$ and $v \otimes 0$ are both equal to the zero vector in $V \otimes W$.

But just as we could get new elements of the direct sum by taking linear combinations of the elements hit by i_V and i_W, there's more to the tensor product than just the elements of the form $v \otimes w$. The elements that can be written this way are called **pure tensors**, but most linear

combinations of pure tensors — like $3v \otimes w - \frac{1}{2}v' \otimes w'$ — won't be pure tensors. In a sense this is another consequence of the fact that p isn't linear: adding two elements of the image of a linear map will keep you in the image, but this is not true of the image of p.

Finally, suppose V and W are finite-dimensional, and we're given bases $v_1, \ldots, v_n$ for V and $w_1, \ldots, w_k$ for W. In this case we can directly build a vector space E satisfying the universal property. Whatever E looks like, it has to contain a vector $v_i \otimes w_j$ for every pair of basis vectors from V and W. But in fact, if we take E to be the nk-dimensional vector space with basis vectors labeled in this way, E will satisfy the universal property. Given a bilinear map $q : V \times W \to A$ for some A, we see that if $q = up$ we have to have $u(v_i \otimes w_j) = q((v_i, w_j))$. But since we said that the $v_i \otimes w_j$'s form a basis, this already uniquely specifies u! To finish the proof, the remaining thing to check is that if we define u in this way we always actually have $u(p((v, w))) = q((v, w))$, even when v and w don't come from the bases we started with. I will leave this last step to you to check.

3.2 Constructing the Tensor Product

In this short section, we'll go through the general construction of tensor products. In the last section we found a basis for the tensor product of two finite-dimensional vector spaces, and therefore in a sense showed that tensor products exist at least in this case. Still, the construction we present here has a couple advantages. In addition to not requiring any assumptions about dimension, it also self-evidently doesn't depend on any choice of basis.

Since the tensor product should come with a bilinear map $p : V \times W \to V \otimes W$, we'll start by building a vector space that has an element for p to hit for every $v \in V$ and $w \in W$. Only after that will we worry about making p bilinear. Specifically, let's define E to be the (gigantic!) vector space with one basis vector for every pair of elements $v \in V$, $w \in W$, which we'll write $[v, w]$. That is, the elements of E are expressions of the form

$$\sum_{i=1}^{r} \lambda_i [v_i, w_i],$$

and addition and scalar multiplication are defined by adding or multiplying these coefficients, so, for example,

$$\left(\sum_{i=1}^{r} \lambda_i [v_i, w_i] \right) + \left(\sum_{i=1}^{r} \mu_i [v_i, w_i] \right) = \sum_{i=1}^{r} (\lambda_i + \mu_i)[v_i, w_i].$$

Since these $[v, w]$'s are a basis for E, there are no linear relations among them whatsoever for different v's and w's. For example, $[2v, w]$ and $2[v, w]$ are different vectors in E, and $[v, 0]$ isn't the zero vector. Because of this, the function sending each $(v, w) \in V \times W$ to $[v, w] \in E$ is very far from being bilinear. To fix this, we'll take a quotient of E to force the required relations to hold. Specifically, let F be the subspace of E spanned by all of the following elements:

- $[v + v', w] - [v, w] - [v', w]$ for each $v, v' \in V$ and $w \in W$,

- $[v, w + w'] - [v, w] - [v, w']$ for each $v \in V$ and $w, w' \in W$,

- $[\lambda v, w] - \lambda[v, w]$ for each $v \in V$, $w \in W$, and $\lambda \in k$,

- $[v, \lambda w] - \lambda[v, w]$ for each $v \in V$, $w \in W$, and $\lambda \in k$.

We then define $V \otimes W = E/F$, and define the map p to take (v, w) to the equivalence class of $[v, w]$ in this quotient; we will write $v \otimes w$ for this equivalence class. We then claim that this construction of $V \otimes W$ satisfies the universal property.

To see this, we first need to see that p is bilinear. But note that the fact that p is bilinear is exactly equivalent to the elements we placed in F being equal to 0. Since we took the quotient by F, this is indeed the case.

Now, suppose we have another bilinear map q from V and W to X. We'd like to find a linear map $u : V \otimes W \to X$ so that $up = q$ and to show that it's unique. Just as we saw when we were discussing direct sums, our choice of u is completely forced: for every element of the form $v \otimes w$, the fact that $up = q$ forces us to say that $u(v \otimes w) = q(v, w)$. But the elements of the form $v \otimes w$ span $V \otimes W$ (since the $[v, w]$'s span E), so if u exists at all there's only one possible choice.

And in fact this choice works. You'll work this out in detail in the exercises, but the sketch is as follows: we can define a linear map $\bar{u} : E \to X$ by sending $[v, w]$ to $q(v, w)$ (and extending the definition to linear combinations of $[v, w]$'s in the obvious way). But $\bar{u}(F) = 0$ — I am leaving this step to you to check — so $\bar{u}$ in fact descends to a map $E/F \to X$.

This completes the proof and the general construction of the tensor product of two vector spaces. Since we are going to be using the tensor product a lot in various constructions that we'll claim are coordinate-free, it is nice to have checked that we can construct it without picking bases for V and W. Despite this, we won't actually have much use for the details of this construction. The power of having a universal property is that, once we know an object exists that satisfies it, the universal property itself is usually all you need going forward.

$\boxed{3.3}$ **Multilinear Maps**

This whole story generalizes quite straightforwardly to give a tensor product of more than two vector spaces. A **multilinear map** from the V_i's to W is a function $m : V_1 \times \cdots \times V_k \to W$ which is linear in each coordinate. (When we want to be specific about the number of vector spaces involved we'll call it a k-**linear map**.) That is, for each index i and each $v_1, \ldots, v_k, v_i' \in V, \lambda \in k$, we have

- $m(v_1, \ldots, v_i + v_i', \ldots, v_k) = m(v_1, \ldots, v_i, \ldots, v_k) + m(v_1, \ldots, v_i', \ldots, v_k)$, and

- $m(v_1, \ldots, \lambda v_i, \ldots, v_k) = \lambda m(v_1, \ldots, v_i, \ldots, v_k)$.

Just as with bilinear maps, we can build a vector space from the V_i's to serve as the universal source of multilinear maps out of them, which we'll call the k-**fold tensor product**, written $V_1 \otimes \cdots \otimes V_k$. That is, there is a multilinear map p from the V_i's to $V_1 \otimes \cdots \otimes V_k$ with the property that, for any other multilinear map m from the V_i's to A, there is a unique linear map $u : V_1 \otimes \cdots \otimes V_k \to A$ for which $m = up$:

$$V_1 \times \cdots \times V_k \overset{p}{\Longrightarrow} V_1 \otimes \cdots \otimes V_k$$

with m and u to A.

This notation raises a couple of questions that it's worth resolving right away. First, as the notation suggests, a vector space satisfying the universal property of $(V_1 \otimes V_2) \otimes V_3$ (about bilinear maps out of $V_1 \otimes V_2$ and V_3) will in fact also satisfy the universal property of $V_1 \otimes V_2 \otimes V_3$ (about trilinear maps out of V_1, V_2, and V_3). You'll show this in the exercises at the end of this section.

Second, there's nothing special about the number three here or about this particular way of arranging the parentheses. There is, for example, an isomorphism between $(V_1 \otimes V_2) \otimes V_3$ and $V_1 \otimes (V_2 \otimes V_3)$, so we don't need to be too careful about where we place the parentheses and we usually won't bother to write them. That is, the two-fold tensor product is "associative" in this way. In fact, one good way to show this is to reduce it to the claim from the preceding paragraph: both ways of arranging the parentheses satisfy the universal property about trilinear maps, and so they must be isomorphic to each other. A similar fact is true about k-linear maps and all possible ways of combining k vector spaces with the two-fold tensor product. This all justifies the use of the notation "$V_1 \otimes \cdots \otimes V_k$" for the k-fold tensor product — if these things weren't true this notation might be ambiguous.

Just as in the bilinear case, we'll often be particularly interested in multilinear maps from k copies of the same vector space V to $\mathbb{R}$. These are called **multilinear** or k-**linear forms on** V. Because of the universal property, a k-linear form on V is the same as a map from $V \otimes \cdots \otimes V$ to $\mathbb{R}$, where there are k copies of V in the tensor product. This tensor product of k copies of V is important enough to have a name: we call it the k-**th tensor power** of V and write it $V^{\otimes k}$. It will be an important part of our discussion of determinants in a later section.

Exercises

1. Show that for any bilinear map $p : V \times W \to X$ and any $v \in V$, $w \in W$, $p(v, 0) = p(0, w) = 0$.

2. As we did with direct sums, show that tensor products are unique, in the following sense: if $p : V \times W \to T$ and $p' : V \times W \to T'$ are two bilinear maps satisfying the universal property defining tensor products, show that there is a unique invertible linear map $\phi : T \to T'$ so that $\phi p = p'$.

3. Finish the missing step in the construction of the tensor product: prove that, given any bilinear map $q : V \times W \to X$, it is always possible to build a map u satisfying the conditions listed there.

4. Use the universal property to show that $V \otimes W$ is isomorphic to $W \otimes V$ for any V, W, and that $U \otimes (V \otimes W)$ is isomorphic to $(U \otimes V) \otimes W$ and to the 3-fold tensor product $U \otimes V \otimes W$.

5. Prove that the universal property of $\mathbb{R} \otimes V$ is satisfied by V, and that the pure tensor $\alpha \otimes v \in \mathbb{R} \otimes V$ corresponds to just $\alpha v \in V$.

6. For vector spaces U, V, W, use the universal property to build an isomorphism between $U \otimes (V \oplus W)$ and $(U \otimes V) \oplus (U \otimes W)$.

7. In one of the exercises to the last section, we found a way to take linear maps $f : V \to V'$ and $g : W \to W'$ and produce a linear map we called $f \oplus g : V \oplus W \to V' \oplus W'$, and we saw that its existence can be deduced just from the universal property of the direct sum.

   (a) Use the universal property of the tensor product in a similar way to produce a linear map $f \otimes g : V \otimes W \to V' \otimes W'$.

   (b) Prove that this construction respects composition in the same way as the direct sum version. That is, if we also have linear maps $f' : V' \to V''$ and $g' : W' \to W''$, then

   $$(f' \otimes g')(f \otimes g) = f'f \otimes g'g.$$

(c) If we have bases for V, W', V', and W', how does the matrix for $f \otimes g$ relate to the matrices for f and g? (The resulting matrix is sometimes called the **Kronecker product** of the f and g matrices.)

4 Duals and Traces

We're now ready to take a look at some familiar constructions from linear algebra, starting with the trace of a linear map. The tools we've built here give enable us to define it in a new way that can serve to illuminate some of its properties.

4.1 Duals, Inner Products, and Transposes

Given any two vector spaces V and W, we can form a new vector space $\mathrm{Hom}(V, W)$ consisting of linear maps from V to W. Each element of $\mathrm{Hom}(V, W)$ is a linear map, and $\mathrm{Hom}(V, W)$ is given the structure of a vector space in the obvious way: given $f, g \in \mathrm{Hom}(V, W)$ and $\lambda \in \mathbb{R}$, $f + g$ and λf are the maps given by $(f + g)(v) = f(v) + g(v)$ and $(\lambda f)(v) = \lambda(f(v))$. There is one particular case which has a special name: we call $\mathrm{Hom}(V, \mathbb{R})$ the **dual of** V, and we write it V^* for short.

When V is finite-dimensional, its dual space V^* has the same dimension. In fact, given a basis $e_1, \ldots, e_n$ of V there is a natural basis we can assign to V^*, which we'll write $\eta_1, \ldots, \eta_n$, given by the rule that $\eta_i(e_j) = 1$ if $i = j$ and 0 otherwise. We call this the **dual basis** of V^*. If we use the e basis for V to write maps from V to $\mathbb{R}$ as $1 \times n$ matrices, then the entries of such a matrix are the coefficients we get when we expand the corresponding element of V^* in terms of the η basis.

It can be tempting to treat V and V^* as though they are basically the same vector space — after all, they have the same dimension, so there is always an isomorphism between them — but it's often not a good idea. The problem isn't that there isn't an isomorphism, it's that there are too many isomorphisms and no good way to choose among them. Even if you wanted to use the existence of the dual basis we just built to specify an isomorphism — sending each e_i to η_i — the resulting map still depends on the original choice of basis. Every vector η_i in the dual basis depends on the *entire* basis we picked for V, not just on the corresponding e_i. (For example, you can check that if V is three-dimensional with basis $\{e_1, e_2, e_3\}$, then replacing this basis with $\{e_1, e_2, e_1 + e_2 + e_3\}$ actually makes you change η_1 and η_2, not η_3.) If you think you will ever care about more than one basis for your vector space, then, keeping the distinction between V and V^* in mind is a good way to avoid becoming confused.

Suppose we have a map $f : V \to W$. Given a map from W to $\mathbb{R}$, we can get a map from V to $\mathbb{R}$ just by composing with f. This gives us a way to build a map we'll call $f^* : W^* \to V^*$ by setting, for any $\eta \in W^*$, $f^*(\eta)(v) = \eta(f(v))$. We'll call f^* the **dual of** f.

What does f^* look like as a matrix after we've picked bases for V and W? Say $e_1, \ldots, e_n$ and $e'_1, \ldots, e'_m$ are bases for V and W respectively, and $\eta_1, \ldots, \eta_n$ and $\eta'_1, \ldots, \eta'_m$ are the dual bases for V^* and W^*. If the entries of f as a matrix are a_{ij}, that means that

$$f(e_j) = \sum_{k=1}^{m} a_{kj} e'_k.$$

If we apply η_i' to both sides, we see that it is equivalent to say that $a_{ij} = \eta_i'(f(e_j))$. Suppose the entries of f^* are b_{ij}, so that

$$f^*(\eta_j') = \sum_{k=1}^{n} b_{kj}\eta_k.$$

We can plug e_i into both sides to see that

$$b_{ij} = f^*(\eta_j')(e_i) = \eta_j'(f(e_i)) = a_{ji}.$$

So the matrix of f^* is just the transpose of the matrix for f — put another way, the dual of a linear map is the coordinate-free version of the transpose of a matrix. This description is missing something, though: it's very common in applications to do things like multiply a matrix by its own transpose, but we seem to have made that impossible by insisting on the distinction between vectors spaces and their duals. Without an isomorphism between V and V^*, what could an expression like AA^T possibly mean?

There is a common setting, though, in which a choice of isomorphism between V and V^* presents itself. Recall that an **inner product** is a bilinear form p on V satisfying three properties:

- It is **symmetric**, meaning that $p(v, w) = p(w, v)$ for every $v, w \in V$.

- It is **positive definite**, meaning that $p(v, v) \geq 0$ for every $v \in V$.

- It is **nondegenerate**, meaning that if we have a v for which $p(v, w) = 0$ for every w, then in fact $v = 0$ (and the same thing with the roles of v and w switched).

Given any bilinear form p on V at all, not necessarily one satisfying these properties, we can use it to construct a linear map $\bar{p} : V \to V^*$: we set $\bar{p}(v)$ to be the map taking w to $p(v, w)$. (That is, $\bar{p}(v)(w) = p(v, w)$.) The bilinearity of p tells us both that $\bar{p}$ is linear as a map from V to V^* and that $\bar{p}(v)$ is linear as a map from V to $\mathbb{R}$. This process is even reversible: from a linear map from V to V^* we can produce a bilinear form on V using the same rule.

This gives us a one-to-one correspondence between bilinear forms on V and maps from V to V^*. If p is nondegenerate, though, we see that $\bar{p}$ is injective: the only way that $\bar{p}(v)$ can be the zero map is if $v = 0$. Since V and V^* have the same dimension, this means $\bar{p}$ is actually an isomorphism! In particular, we see that choosing an inner product on V gives us a way to identity V and V^*. (The identification still depends on your original choice of inner product, but not on a choice of basis.)

But if you do have an inner product in mind for both V and W and you use it to identify each of those vector spaces with their duals, you are then free to think of f^* as a map from W to V. There is even an equivalent way to think about f^* that refers more directly to the inner product: given a map $f : V \to W$ and inner products p on V and q on W, the **adjoint** of f is the unique map $f^* : W \to V$ for which

$$q(f(v), w) = p(v, f^*(w))$$

for any vectors $v \in V$ and $w \in W$. In the exercises you will check that there always is such a unique map f^* and that it's the same map you get by taking the dual of f and turning it into a map from W to V in the way described above (hence the use of the same notation).

The upshot of all this is the following: any time you see a transpose of a matrix, it is probably a sign that either you are actually looking at a map between dual spaces or, more likely, that you have already chosen an inner product and you're actually looking at an adjoint. One good example of the former is the well-known fact that a matrix has rank 1 if and only if it can be written as wv^T for nonzero vectors v and w. You'll explore this in one of the exercises.

When the transpose is actually an adjoint, there are some settings in which there might be more than one inner product you could be using, and when that's true it's it probably doesn't make sense to transpose your matrix without being very careful about which of these things you actually mean and which inner product you are using.

4.2 The Trace Without Coordinates

The coordinate-free definition of the trace comes out of some of the constructions we just described. Our goal will be to build a linear map $\text{tr} : \text{Hom}(V, V) \to \mathbb{R}$. The construction of tr will in turn depend on a fact about Hom: when V and W are finite-dimensional, there is an isomorphism between $V^* \otimes W$ and $\text{Hom}(V, W)$. We'll prove this now.

We start by building a map $h : V^* \otimes W \to \text{Hom}(V, W)$, which by the universal property of the tensor product is the same as building a bilinear map from V^* and W to $\text{Hom}(V, W)$. (This map will in fact always exist, but it will only be an isomorphism if V and W are finite-dimensional.) To do this, given $\eta \in V^*$ and $w \in W$, we need to construct an element of $\text{Hom}(V, W)$, that is, linear map from V to W. The definition of this linear map is simple: we send any $v \in V$ to $\eta(v)w$ — recall that $\eta \in V^*$, so $\eta(v)$ is a scalar. I encourage you to verify that this is indeed a bilinear map (that is, it's linear in η and w separately).

We'll start our proof that h is an isomorphism by showing that it's injective. An arbitrary element of $V^* \otimes W$ looks like $\sum_i \eta_i \otimes w_i$, with $\eta_i \in V^*$, and $w_i \in W$. Moreover, we can assume that all the w_i's that appear in this sum are linearly independent. To see this, suppose that some $w_j = \sum_i \lambda_i w_i$. We can use this fact to rewrite the sum so that w_j doesn't appear in it anymore:

$$\sum_i \eta_i \otimes w_i = \sum_{i \neq j} (\eta_i \otimes w_i) + \eta_j \otimes \left(\sum_{i \neq j} \lambda_i w_i \right) = \sum_{i \neq j} (\eta_i + \lambda_i \eta_j) \otimes w_i.$$

This gives us a new way of writing the same element of $V^* \otimes W$ with one fewer term in the sum. We may keep doing this until the remaining w_i's are linearly independent.

So now suppose we have some nonzero $t = \sum_i \eta_i \otimes w_i$, where all the w_i's are linearly independent and no $\eta_i = 0$. (If some $\eta_i = 0$ we can just remove that term from the sum.) Since, say, $\eta_1 \neq 0$, we can pick some $v \in V$ for which $\eta_1(v) \neq 0$. Then consider $h(t)(v) = \sum_i \eta_i(v)w_i$. This sum can't be zero: the w_i's are all linearly independent and we know that at least one coefficient, $\eta_1(v)$, is nonzero.

This means that $h(t)$ isn't the zero map, that is, it isn't the zero vector in $\text{Hom}(V, W)$. Since this was true for every nonzero $t \in V^* \otimes W$, this means h is injective.

Finally, suppose $\dim V = n$ and $\dim W = k$. Then both $V^* \otimes W$ and $\text{Hom}(V, W)$ have dimension nk. (Once we've picked bases for V and W, elements of $\text{Hom}(V, W)$ can be written as $k \times n$ matrices.) So, if V and W are finite-dimensional, h is an injective linear map between two vector spaces of the same dimension. This means it is an isomorphism, finishing the proof.

With this isomorphism in hand we're ready to construct the trace. Take a map $f : V \to V$. Thinking of f as an element of $\text{Hom}(V, V)$, we can use the inverse of the map we just built to get an element of $V^* \otimes V$. So now we just need a way to turn this into a scalar, that is, we need a map $c : V^* \otimes V \to \mathbb{R}$. But there is a very natural bilinear map from V^* and V to $\mathbb{R}$, namely the one that takes η and v to $\eta(v)$. (This is often called the **contraction** map.) So this lands us in $\mathbb{R}$, completing the construction of the trace. The whole thing is summarized in the following diagram:

$$\mathrm{Hom}(V,V) \xrightarrow{\ h^{-1}\ } V^* \otimes V \xrightarrow{\ c\ } \mathbb{R}$$
$$\underbrace{\hspace{5cm}}_{\mathrm{tr}}$$

Let's see what this looks like when we pick a basis $v_1, \ldots, v_n$ for V. Write $\eta_1, \ldots, \eta_n$ for the dual basis of V^*. Then applying the definition of h, we see that $h(\eta_j \otimes v_i)$ is the linear map which takes v_j to v_i and takes the rest of the basis to 0.

Having picked a basis for V, we can write linear maps from V to V as matrices, and you will prove in the exercises the matrix for $h(\eta_j \otimes v_i)$ is the one with a 1 in its (i, j) entry and zeroes everywhere else. So the coefficients we see when we expand $\mathrm{Hom}(V, V)$ in this basis are just the entries of the matrix: if $f : V \to V$ has entries a_{ij}, then what we've shown here is that

$$f = \sum_{i,j} a_{ij} h(\eta_j \otimes v_i).$$

So let's use this to compute the trace of f according to our new definition. First, we just showed that

$$h^{-1}(f) = \sum_{i,j} a_{ij} \eta_j \otimes v_i,$$

so we just need to apply c to this. By our definition, though, $c(\eta_j \otimes v_i) = \eta_j(v_i)$, which is 1 if $i = j$ and 0 if $i \neq j$. So the only terms in the sum to survive are the ones in which $i = j$, and we arrive at the final answer:

$$\mathrm{tr}\, f = c(h^{-1}(f)) = \sum_{i,j} a_{ij} c(\eta_j \otimes v_i) = \sum_i a_{ii}.$$

We have recovered the usual formula for the trace as the sum of the diagonal entries of the matrix. While it doesn't take too long to prove directly that a change of basis preserves the sum of diagonal entries of a matrix, the usual proof of this fact can make it difficult to gain insight into why you should have expected something like this to be true. But in defining the trace in terms of c as we did here, we never picked a basis at all, so we get this fact "for free" — this definition communicates the *reason* the trace is independent of the choice of basis, rather than just observing that the answer somehow comes out the same when you manipulate the indices in a sum the right way.

Exercises

1. (a) Prove the existence and uniqueness of adjoints: given a linear map $f : V \to W$ and inner products p on V and q on W, there is a unique map $f^* : W \to V$ for which

$$q(f(v), w) = p(v, f^*(w))$$

for any $v \in V$ and $w \in W$.

   (b) Prove that if we use p and q to identify V and W with their duals, this identifies the adjoint of f with the dual of f.

2. Construct a linear map from V to $(V^*)^*$ without referring to bases, and prove that it's an isomorphism when V is finite-dimensional. If we then pick a basis $e_1, \ldots, e_n$ for V and consider the dual basis $\eta_1, \ldots, \eta_n$ of V^*, prove that your isomorphism identifies the dual basis for $(V^*)^*$ with the original basis for V.

3. Prove that the matrix for $h(\eta_j \otimes v_i)$ is the one with a 1 in its (i, j) entry and zeroes everywhere else.

4. For this problem, consider two linear maps $f : U \to V$ and $g : V \to W$ where U, V, and W are all finite-dimensional.

   (a) We can think of f as living in $\mathrm{Hom}(U, V)$ and g as living in $\mathrm{Hom}(V, W)$. Any pair of linear maps like this can be composed to give $gf \in \mathrm{Hom}(U, W)$. Prove that composition is bilinear. Composition therefore gives a linear map we'll call

   $$k : \mathrm{Hom}(U, V) \otimes \mathrm{Hom}(V, W) \to \mathrm{Hom}(U, W),$$

   so that $gf = k(f \otimes g)$.

   (b) We can apply our isomorphism h between $\mathrm{Hom}(V, W)$ and $V^* \otimes W$ to all three of these Hom spaces and turn k into a map

   $$k' : U^* \otimes V \otimes V^* \otimes W \to U^* \otimes W.$$

   Recall that in the last section we defined a map $c : V^* \otimes V \to \mathbb{R}$ by setting $c(\eta \otimes v) = \eta(v)$. Verify that, in the same way, we can construct a multilinear map m from U^*, V, V^*, and W to $U^* \otimes W$ which "just contracts the V and V^* parts," that is,

   $$m(\phi, v, \eta, w) = \eta(v) \cdot (\phi \otimes w),$$

   which by the universal property produces a linear map

   $$c' : U^* \otimes V \otimes V^* \otimes W \to U^* \otimes W.$$

   Prove that c' and k' are in fact the same linear map. [Hint: Since you already know they're both linear and the pure tensors span, it's enough to check this equality on pure tensors.]

   (c) Suppose we pick bases for U, V, and W. What does the equality of k' and c' mean when you write everything out in coordinates?

5. Now consider two linear maps $f : V \to W$ and $g : W \to V$.

   (a) As in the previous problem, we can use h to think of $f \otimes g$ as living in $V^* \otimes W \otimes W^* \otimes V$. Like we did with m and c' in the last section, build a linear map

   $$j : V^* \otimes W \otimes W^* \otimes V \to \mathbb{R}$$

   which contracts both the V's and the W's. Work out what this map is in detail and show that $j(f \otimes g) = \mathrm{tr}(gf)$.

   (b) Use this to prove that $\mathrm{tr}(fg) = \mathrm{tr}(gf)$. What does your proof look like in coordinates if you pick bases?

6. Prove that a linear map $f : V \to W$ has rank 1 if and only if, for some nonzero $\eta \in V^*$ and $w \in W$, we have $f = h(\eta \otimes w)$. [Hint: If f has rank 1, then there is a basis of V for which f takes the first basis vector to something nonzero and the rest of the basis to zero.]

7. In coordinates, the fact from the previous problem is often stated in the following form: f has rank 1 if and only if, for nonzero vectors $v \in V$ and $w \in W$, we have $f = wv^T$. This problem is about connecting these two formulations.

(a) In coordinates, it is often useful to think of a vector in an n-dimensional vector space as an $n \times 1$ matrix. In our coordinate-free language, this corresponds to finding an isomorphism between V and $\mathrm{Hom}(\mathbb{R}, V)$. What is this isomorphism?

(b) There are two ways we could make sense of the notation "v^T" corresponding to the two interpretations of the transpose we discussed in this section. Show that if we think of it as a dual we end up with an element of $(V^*)^*$, and that this results in the same isomorphism between V and $(V^*)^*$ that you found in an earlier exercise. [Hint: For the first part, note that you have a coordinate-free isomorphism between $\mathbb{R}^*$ and $\mathbb{R}$.]

(c) Suppose we have an inner product q on V. Prove that, if you think of transposes as adjoints, v^T corresponds to the map from $g : V \to \mathbb{R}$ given by $g(w) = q(v, w)$.

(d) Which of these two interpretations is at work when we write a rank-1 matrix as wv^T? Prove that this matrix multiplication produces the same map from V to W as the $h(\eta \otimes w)$ construction you used in the problem about rank-1 maps. [Hint: You can prove this directly, but there is also an argument that uses the problem above about composing linear maps.]

5   The Determinant Without Coordinates

We are now ready for the coordinate-free treatment of the determinant. The determinant of a linear map measures how it affects volumes, so we'll start by seeing how to talk about volumes using the tools we've built.

5.1   The Determinant in Two Dimensions

We'll first go through the construction of the determinant in the two-dimensional case, where some aspects are simpler. Suppose V is two-dimensional, and consider two vectors $v, w \in V$. The **parallelogram spanned by v and w** is the set

$$\{av + bw : 0 \le a \le 1, 0 \le b \le 1\}.$$

It looks like this:
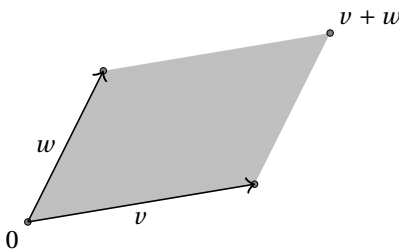
The key thing to notice is that the area of such a parallelogram behaves like a bilinear function of v and w. For example, writing $A(v, w)$ for the area of the parallelogram defined by v and w, we must have

$$A(v + v', w) = A(v, w) + A(v', w).$$

One way to see this is to notice that the triangle cut out by the top three points is congruent to the one cut out by the bottom three points in the following picture:
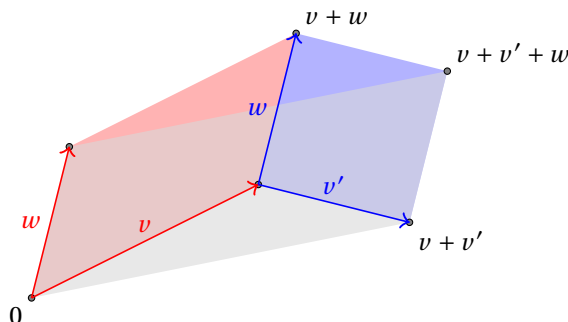
You can draw a very similar picture for multiplying by a positive scalar.

The fact that A is bilinear has a somewhat counterintuitive consequence: if A is always going to be bilinear, then, switching the roles of v and $v + v'$ in the equation from the last paragraph, we also require that

$$A(v, w) = A(v + v', w) + A(-v', w).$$

This forces us to set $A(-v', w) = -A(v', w)$; if they're not both zero, then one of these two numbers has to be negative.

In fact, there's yet another reason that A will sometimes have to take on negative values. If you plug the same vector in for both arguments to A, you should get zero: the resulting "parallelogram" is just a line segment, which has no area. That is, $A(v, v) = 0$ for each v. But if we plug a sum of two vectors into this formula, we see that

$$0 = A(v + w, v + w) = A(v + w, v) + A(v + w, w) = A(v, v) + A(w, v) + A(v, w) + A(w, w).$$

Using this fact again, we can cancel two of the four terms from the right side of this equation, giving us that $A(v, w) = -A(w, v)$. Since this is true for arbitrary vectors $v, w \in V$, we see that, unless $A(v, w) = 0$, one of $A(v, w)$ or $A(w, v)$ has to be negative.

This gives us a way of thinking about what A represents. If A is indeed a bilinear form, then it is not quite the area of the parallelogram, but rather a "signed area"; the absolute value $|A(v, w)|$ is the area, and the sign tells us about the **orientation** of v and w. There are many ways to think about what orientation is telling you about v and w, and a full discussion of this is worth a whole article in itself. One useful picture in this two-dimensional setting is to imagine rotating around the origin in the plane starting at v and moving toward at w; the orientation then tells you whether the direction that gets you to w the fastest is clockwise or counter-clockwise. Whether you use this picture or any other, though, you can reverse the orientation, by switching v and w, by replacing v with $-v$, or by replacing w with $-w$.

A bilinear map q for which $q(v, w) = -q(w, v)$ is called an **alternating bilinear map**. The preceding discussion suggests a way we might bring areas, and therefore the determinant, into the framework we've built in this article: we would like an object which has the same relationship to alternating bilinear maps that the tensor product has to all bilinear maps. We'll see how to do this now.

Let q be a bilinear map out of V, and consider the corresponding map $p : V \otimes V \to W$. Saying q is alternating is equivalent to saying that, for every $v, w \in V$, $p(v \otimes w) = -p(w \otimes v)$, or equivalently that $p(v \otimes w + w \otimes v) = 0$. So, letting $D \subseteq V \otimes V$ be the subspace spanned by all elements of the form $(v \otimes w + w \otimes v)$, q is alternating if and only if p takes D to zero.

Recall that in the exercises in the first section we saw that there is a one-to-one correspondence between maps out of a vector space E that are zero on a subspace E' and maps out

of the quotient E/E'. Putting this all together, then, we get the object we should use as the universal source of alternating bilinear maps: the quotient $(V \otimes V)/D$. This vector space is called the **second exterior power** of V, and it's written $\Lambda^2 V$. An area function like A, then, can be represented by a linear map from $\Lambda^2 V$ to $\mathbb{R}$.

Although we started this discussion by considering the case where V is two-dimensional, the construction of $\Lambda^2 V$ makes sense for any vector space. For $v, v' \in V$, we'll write $v \wedge v'$ for the element of $\Lambda^2 V$ corresponding to the pure tensor $v \otimes v' \in V \otimes V$; these elements are called **pure wedges**. (The symbol $\wedge$ is often pronounced "wedge.") The function $(v, v') \mapsto v \wedge v'$ is an alternating bilinear map, which implies some linear relations among these wedges. We have the bilinear relations like $(v + w) \wedge v' = v \wedge v' + w \wedge v'$ that were also true of the tensor product, but also some new ones, like $v \wedge v' = -v' \wedge v$ and $v \wedge v = 0$.

Just as it is often useful to think of a vector as specifying a direction and an amount of length along that direction, you can think of $v \wedge v'$ as specifying a plane — the one spanned by v and v' — and an amount of (oriented) area in that plane. You can even picture the parallelogram spanned by v and v', as long as you remember that you will get the same element of $\Lambda^2 V$ from any other parallelogram in the same plane with the same orientation and area. Just as v and $-v$ point in opposite directions along the same line, $v \wedge v'$ and $-v \wedge v' = v' \wedge v$ are pieces of area with opposite orientations. Just as in the tensor product, though, not every element of $\Lambda^2 V$ is a pure wedge; a general element is a linear combination of pure wedges.

What does this have to do with determinants? If V is a two-dimensional vector space and $f : V \to V$ is a linear map, then the determinant of f tells us what f does to areas: applying f multiplies areas by a factor of $|\det f|$, and f is orientation-preserving if and only if $\det f$ is positive. We just introduced $\Lambda^2 V$ as a way of talking about pieces of area in V, so this suggests we should try to find a way to use f to produce a linear map from $\Lambda^2 V$ to itself.

It is in fact possible to do this — and we can construct the map from the universal property using the same technique we used for the direct sum and the tensor product in the exercises. Given any linear map $f : V \to W$, we're looking for a map we'll call $\Lambda^2 f : \Lambda^2 V \to \Lambda^2 W$. To do this, it is enough to construct an alternating bilinear map q from V to $\Lambda^2 W$. But we can simply define $q(v, v') = f(v) \wedge f(v')$; I encourage you to check that this is indeed an alternating bilinear map. Just as in the tensor product case, this construction tells us what $\Lambda^2 f$ does to pure wedges: $(\Lambda^2 f)(v \wedge v') = f(v) \wedge f(v')$. I leave it to you to check that, if we also have a linear map $g : W \to X$, then $\Lambda^2(gf) = (\Lambda^2 g)(\Lambda^2 f)$; it is also possible to deduce this just from the universal property.

We will be most interested in the case where f goes from V to itself, giving us $\Lambda^2 f : \Lambda^2 V \to \Lambda^2 V$. If we're going to define the determinant as "the factor f multiplies areas by" then $\Lambda^2 f$ ought to be a scalar. And in fact it is: we will show momentarily that when V is two-dimensional, then $\Lambda^2 V$ is one-dimensional, so *every* linear map on it is a scalar. (This should fit with the geometric intuition for $\Lambda^2 V$ discussed earlier: when V is two-dimensional there's only one plane for the area element to live in, so all that matters is the area and orientation of the parallelogram.)

Let's verify this. Pick any basis $\{e_1, e_2\}$ of V. Recall that then $\{e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2\}$ gives a basis for $V \otimes V$, which means that their images in $\Lambda^2 V = (V \otimes V)/D$ have to span. But $e_1 \wedge e_1$ and $e_2 \wedge e_2$ are both zero, and $e_2 \wedge e_1 = -e_1 \wedge e_2$, so $e_1 \wedge e_2$ actually spans $\Lambda^2 V$ all by itself.

This proves that $\Lambda^2 V$ is *at most* one-dimensional. To show it's at least one-dimensional and finish the proof, it's enough to show that there's a single nonzero vector in it. (There are probably simpler ways to do this than what we are about to do, but I have chosen this proof because it is the one that generalizes to the higher-dimensional case.) In fact, $e_1 \wedge e_2$ is nonzero, and we'll show this by finding a linear map $\Lambda^2 V \to \mathbb{R}$ which takes it to 1. By the universal property that started us off, this is the same as finding an alternating bilinear form on V that takes the pair

(e_1, e_2) to 1.

Let $\{\eta_1, \eta_2\}$ be the dual basis on V^* — that is, they're the linear functions $V \to \mathbb{R}$ given by $\eta_1(e_1) = 1$, $\eta_1(e_2) = 0$, and vice versa for η_2. Then our alternating bilinear form is:

$$q(v, v') = \eta_1(v)\eta_2(v') - \eta_1(v')\eta_2(v).$$

I leave it to you to check that this is indeed an alternating bilinear form and that $q(e_1, e_2) = 1$.

So we have our recipe for the determinant of a linear map from a two-dimensional vector space to itself. Given $f : V \to V$, produce the map $\Lambda^2 f : \Lambda^2 V \to \Lambda^2 V$. Since $\Lambda^2 V$ is one-dimensional, this second map has to just be multiplication by a unique scalar, and that scalar is what we call the determinant.

An important thing to keep in mind is that it's not possible to attach a numerical area to an element of $\Lambda^2 V$ without any other information. Declaring that, say, the unit square has area 1 requires knowing which parallelogram is the unit square, and that only means something after you have picked a basis. But in spite of this, the fact that $\Lambda^2 V$ is one-dimensional means that every nonzero element is a scalar multiple of every other, so it is meaningful to ask *how many times larger* one element is than another. So we can define the determinant, and even interpret it is "the factor areas are multiplied by," without a basis for V. For example, if f is the map defined by multiplication by 2, then

$$(\Lambda^2 f)(v \wedge v') = (2v) \wedge (2v') = 4(v \wedge v'),$$

and so we can conclude that $\det f = 4$, even without having made the decisions that assign an area to the parallelogram spanned by v and v'.

In the same way, we cannot declare whether an element of $\Lambda^2 V$ is positively or negatively oriented without making additional choices, but we can ask whether two such (nonzero) elements have the same sign, since this is the same question as whether one is a positive or negative scalar multiple of the other. So, just as for areas, it is sensible to ask whether f preserves or reverses orientations even without having made any other decisions about V.

We are about to see how to generalize all this to larger dimensions, but before moving on I strongly encourage you to check that it's possible to extract the usual definition of the determinant from all this: if we've picked a basis for V and f is given by the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in that basis, then $\det f = ad - bc$.

5.2 The Determinant in General

We can generalize this whole story from areas to volumes pretty straightforwardly. To talk about k-dimensional volumes rather than areas we will need something like a bilinear form but that takes k vectors instead of two.

We'll say that a multilinear form m on a vector space V is alternating if switching two adjacent arguments to m switches the sign. In symbols, this means that for every $v_1, \ldots, v_k \in V$, we have

$$m(v_1, \ldots, v_i, v_{i+1}, \ldots, v_k) = -m(v_1, \ldots, v_{i+1}, v_i, \ldots, v_k).$$

Define $D \subseteq V^{\otimes k}$ to be the subspace cut out by all elements of the form

$$(v_1 \otimes \cdots \otimes v_i \otimes v_{i+1} \otimes \cdots \otimes v_k) + (v_1 \otimes \cdots \otimes v_{i+1} \otimes v_i \otimes \cdots \otimes v_k).$$

For exactly the same reason as in the bilinear case, the alternating multilinear forms correspond exactly to the maps $V^{\otimes k}$ taking D to zero, so we give a name to the quotient $(V^{\otimes k})/D$: we call

it the k**'th exterior power of** V and write it $\Lambda^k V$. For exactly the same reason as for Λ^2, given a linear map $f : V \to W$ we can produce a linear map called $\Lambda^k f : \Lambda^k V \to \Lambda^k W$, and this construction has the same relationship to composition as Λ^2, that is, $\Lambda^k(gf) = (\Lambda^k g)(\Lambda^k f)$.

As before, the image of the pure tensor $v_1 \otimes \cdots \otimes v_k$ is called a "pure wedge," written $v_1 \wedge \cdots \wedge v_k$. The definition of D gives us some relations among these pure wedges, like

$$v_1 \wedge \cdots \wedge v_{i+1} \wedge v_i \wedge \cdots \wedge v_k = -v_1 \wedge \cdots \wedge v_i \wedge v_{i+1} \wedge \cdots \wedge v_k$$

and

$$v_1 \wedge \cdots \wedge v_i \wedge v_i \wedge \cdots \wedge v_k = 0.$$

We thought of elements of $\Lambda^2 V$ as representing parallelogram-shaped pieces of area, and in the same way we can think of elements of $\Lambda^k V$ as representing a piece of k-dimensional volume. The **parallelotope** spanned by a set of vectors $v_1, \ldots, v_k$ is the set of all vectors of the form $\alpha_1 v_1 + \cdots + \alpha_k v_k$ where each α_i is between 0 and 1:
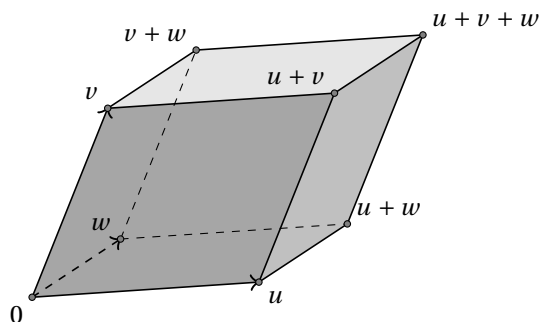
Parallelotopes are exactly the shapes you get when you apply an invertible linear map to a hypercube, and they will serve as the higher-dimensional analogues of the parallelograms we used when discussing Λ^2.

The definition of an alternating multilinear form tells you that you introduce a minus sign when you switch two adjacent arguments. This means that when you perform a more complicated reordering of the arguments, it might be hard to tell whether the resulting sign should be positive or negative; it depends on whether your reordering results from an even or odd number of such switches. It's not even immediately clear that this is well-defined: why can't you get the same reordering in two different ways, one using an odd number or switches and another using an even number? For this and other reasons, it will be helpful to be able to figure out this sign simply by looking at the reordering itself.

A **permutation of** k is a way of reordering the numbers $1, 2, \ldots, k$, that is, a bijection from the set $\{1, 2, \ldots, k\}$ to itself. Given such a permutation π, we'll say an **inversion of** π is any pair of numbers whose order is reversed by π — any pair $i < j$ where $\pi(i) > \pi(j)$. If π has an odd number of inversions, we'll say that π is an **odd permutation**; otherwise, we'll call it **even**. The oddness or evenness of a permutation is usually referred to as its **sign**.

Suppose, for example, that π is the permutation of 5 whose values are, in order, 3, 1, 2, 5, 4. (That is, $\pi(1) = 3$, $\pi(2) = 1$, and so on.) We then have an inversion for every pair of numbers in this list, not necessarily next to each other, where the larger number comes first. There are three of these: $(3, 1)$, $(3, 2)$, and $(5, 4)$, so π is an odd permutation.

In the exercises, you'll show that if you take an odd permutation and switch two adjacent elements, you get an even permutation, and vice versa, and that this implies that no matter how

you break up a permutation into such switches, the sign will always be the same. You may be able to change the *number* of switches — for example, you could always add two switches by doing the same switch twice in a row — but for an odd permutation this number will always be odd and for an even permutation it will always be even. This means we have an answer to our question from before: if you reorder the arguments to an alternating multilinear form, you get have to multiply by −1 if the permutation you used is odd and you get the same result if it's even.

When V is a k-dimensional vector space we'll use all this to define the determinant of a linear map from V to itself. By analogy with the two-dimensional version, we should prove that in this case $\Lambda^k V$ has dimension 1. In fact, we'll prove something more general: if $\dim V = n$, then $\dim(\Lambda^k V) = \binom{n}{k}$.

We'll prove this now. Pick a basis $e_1, \ldots, e_n$. Then I claim we get a basis for $\Lambda^k V$ by taking all possible ways of wedging together k basis vectors in increasing order, that is, all vectors of the form

$$e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k}$$

for sequences $1 \le i_1 < i_2 < \cdots < i_k \le n$. Since there are $\binom{n}{k}$ of these sequences, this will prove the result.

As in the two-dimensional case, all the k-fold tensor products of the e_i's form a basis for $V^{\otimes k}$, and since $\Lambda^k V$ is a quotient of $V^{\otimes k}$, their images span $\Lambda^k V$. Given such an element $e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k}$, we can reorder the e_{i_j}'s until they appear in increasing order; the result will be the same element of $\Lambda^k V$, except possibly with an extra minus sign. If, after doing this, two neighboring i_j's are equal, then the result is zero in $\Lambda^k V$.

So we have in fact shrunk our spanning set to only include the pure wedges we are after, and it remains to show that they are linearly independent. To do this, we will construct, for each increasing sequence of indices, a linear function $\Lambda^k V \to \mathbb{R}$ which is 1 on the corresponding pure wedge and 0 on all the others. (You should verify to yourself now that this is enough to prove the linear independence we're looking for.) By the universal property, given some sequence $i_1, \ldots, i_k$, we want to construct an alternating k-linear form q on V so that $q(e_{i'_1}, e_{i'_2}, \ldots, e_{i'_k})$ is 1 if $i'_1, \ldots, i'_k$ is exactly the same sequence as $i_1, \ldots, i_k$, but is 0 for all other increasing sequences of indices. (Note that we only need this last property for other *increasing* sequences, since these are the elements whose linear independence we are trying to establish. Indeed, the fact that q is alternating means that any reordering of our sequence will give another sequence on which q is nonzero, but none of these sequences is increasing.)

For ease of reading, we'll do this just for the sequence $1, 2, \ldots, k$. (It is a useful exercise to show that you can build the alternating bilinear form for any other sequence out of this one.) The method will directly generalize what we did in the two-dimensional case. Write $\{\eta_1, \eta_2, \ldots, \eta_n\}$ for the dual basis on V^*. If we didn't require q to be alternating, we could accomplish our goal by setting

$$q(v_1, \ldots, v_k) = \eta_1(v_1) \cdot \eta_2(v_2) \cdots \eta_k(v_k) = \prod_{i=1}^{k} \eta_i(v_i).$$

This does the right thing to our basis vectors but isn't alternating, so we need to fix it. Given a permutation π, we'll write $\operatorname{sgn}(\pi)$ to mean 1 if π is even and −1 if π is odd. Then we'll define

$$q(v_1, \ldots, v_k) = \sum_{\pi} \operatorname{sgn}(\pi) \prod_{i=1}^{k} \eta_i(v_{\pi(i)}).$$

For example, if $k = 3$ this works out to be:

$$q(v_1, v_2, v_3) = \eta_1(v_1)\eta_2(v_2)\eta_3(v_3) + \eta_1(v_2)\eta_2(v_3)\eta_3(v_1) + \eta_1(v_3)\eta_2(v_1)\eta_3(v_2)$$
$$-\eta_1(v_2)\eta_2(v_1)\eta_3(v_3) - \eta_1(v_1)\eta_2(v_3)\eta_3(v_2) - \eta_1(v_3)\eta_2(v_2)\eta_3(v_1).$$

Write πs_i for the permutation that first switches i and $i + 1$ and then performs π. Switching the i'th and $(i+1)$'st arguments to q has the effect of replacing the term corresponding to π with the one corresponding to πs_i, except that the sign in front is still $\text{sgn}(\pi)$. But, as we mentioned above, you will prove on the exercises that π and πs_i always have opposite signs. So switching the i'th and $(i+1)$'st arguments to q in fact just has the effect of reordering all the terms of the sum and switching the sign on each, and this proves that q is alternating.

If you plug some sequence of our basis vectors into q, the only way a term of the sum can be nonzero is if all the indices on the η_i's match up exactly with the indices on the e_j's; a single mismatch means the whole product is zero. If the sequence of basis vectors being plugged in isn't the one we used to build q, then such a match will never happen and the whole sum is zero. If it is the sequence we used to build q, then the only nonzero term of the sum will be 1, which was our goal, finishing the proof.

Now that this has been established, we can finish up the same way we did in the two-dimensional case. Given a map $f : V \to V$ where V is n-dimensional, we can take its n'th exterior power $\Lambda^n f : \Lambda^n V \to \Lambda^n V$. Since $\Lambda^n V$ has dimension $\binom{n}{n} = 1$, every such map is just multiplication by scalar. We call that scalar the determinant of f.

As in the two-dimensional case, the sign of the determinant tells you whether or not f reverses orientations. It is a bit more difficult to form a good mental picture of orientation of an n-tuple of vectors when n is bigger than 2. (One picture, which may or may not be all that satisfying, is that two n-tuples of vectors have the same orientation if it is possible to continuously transform one into the other while keeping them linearly independent the whole time.) But the essential algebraic features survive from the two-dimensional case. As was true there, it doesn't make sense to say that an element of $\Lambda^n V$ is "positively oriented" without having made any other choice, but we can still ask whether two such elements have the same sign, and it therefore is possible to decide whether an invertible linear map is orientation-preserving or orientation-reversing.

If we pick a basis for V and write f as a matrix, a formula for the determinant also comes out of this perspective pretty directly. Say the entries of f are a_{ij}, so that $f(e_j) = \sum_i a_{ij} e_i$ for each j. We can then see what $\Lambda^n f$ does by applying it to any nonzero element of $\Lambda^n V$, say the pure wedge $e_1 \wedge \cdots \wedge e_n$. (We know this element is nonzero because it's sent to 1 by the map we constructed in the proof!)

This is sent by $\Lambda^n f$ to

$$f(e_1) \wedge \cdots \wedge f(e_n) = \left(\sum_i a_{i1} e_i\right) \wedge \cdots \wedge \left(\sum_i a_{in} e_i\right).$$

Since $\wedge$ is multilinear, we can collect all the terms into one giant sum. This sum will have n^n terms, one for each ordered sequence of n basis vectors; the term corresponding to the sequence $e_{i_1}, e_{i_2}, \dots, e_{i_n}$ will look like

$$(a_{i_1,1} a_{i_2,2} \cdots a_{i_n,n})(e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_n}).$$

If any basis vector appears twice in one of these wedges, though, that term will be zero. So we in fact have only the terms in which each basis vector appears exactly once, and each of

these corresponds to a permutation of n. And if we take the term where the basis vectors are permuted by some permutation π, we can reorder the vectors in the wedge to make it look like $e_1 \wedge \cdots \wedge e_n$ as long as we multiply by the sign of π. Putting this all together, we have

$$(\Lambda^n f)(e_1 \wedge \cdots \wedge e_n) = \left(\sum_\pi \mathrm{sgn}(\pi) \prod_{j=1}^n a_{\pi(j),j} \right) e_1 \wedge \cdots \wedge e_n,$$

and therefore that sum in parentheses is the determinant of f.

Defining the determinant this way in terms of the exterior power has a couple advantages. First, since we didn't mention bases at all in the definition, we get "for free" that the determinant is independent of the choice of basis. We can also use the fact that Λ^k respects composition — that is, $\Lambda^k(fg) = (\Lambda^k f)(\Lambda^k g)$ — to deduce that $\det(fg) = \det f \cdot \det g$, since composing two scalar maps just amounts to multiplying the corresponding scalars. And in the exercises you will investigate the relationship between exterior powers and duals to see that the determinant of a matrix is the same as the determinant of its transpose.

But, while they are nice to know about, I don't think any of these things is really the point of this whole endeavor. It's possible (if maybe cumbersome) to prove each of them directly from the sum-over-permutations definition of the determinant. Rather, I see more conceptual benefits to taking the coordinate-free perspective on the determinant — and on everything else we've discussed in this long article. When I learned about traces and determinants for the first time, they seemed like arbitrary manipulations performed on the entries of a matrix, and the fact that, say, the determinant can tell me whether a matrix is invertible seemed like magic. My hope is that this new way of looking at algebra can help some readers feel as though there is some order and reason behind all these formulas or that it does something to clarify when some construction depends on making a choice of a basis or inner product.

When I first learned the things in this article, I found them very satisfying; it felt to me that a lot more of the mathematical universe made sense than before. (Plus I think it's a lot of fun.) But even this isn't really the reason mathematicians invented category theory, which is where much of these definitions come from. While the coordinate-free perspective on linear algebra is nice, all of these concepts were developed well before it came along. If you go on to study more mathematics, though, you'll find that there are direct analogues to all of these concepts for more complicated mathematical objects, like representations of a group, modules over a ring, or vector bundles on a topological space. In these settings it's often not possible to do anything as straightforward as picking a basis of a vector space, and it therefore becomes much more important to keep careful track of the extent to which the constructions you use depend on arbitrary choices. Once this is true, the tools of category theory very quickly go from being helpful but inessential to being completely indispensable.

Exercises

1. In this problem we will examine the basic properties of the sign of a permutation, in particular proving enough about it for the purposes for which it was employed in this article. Recall the definition we started with: an even (or odd) permutation is one with an even (or odd) number of inversions.

   Throughout this problem we will think of permutations as functions from $\{1, 2, \ldots, n\}$ to itself. In particular, they can be composed, and we'll write composition multiplicatively, so if π and σ are permutations, then $\pi\sigma$ takes i to $\pi(\sigma(i))$.

(a) Write s_i for the permutation that switches i and $i + 1$. These are called **neighbor transpositions**. It is often helpful to write down a permutation π by listing the numbers $\pi(1), \pi(2), \dots, \pi(n)$ in order. Given such a list for π, what does the list for πs_i look like? What about $s_i \pi$?

(b) Prove that the number of inversions of πs_i is either one more or one less than the number of inversions of π, and therefore πs_i is odd if π is even, and vice versa.

(c) Any permutation can be built up from compositions of neighbor transpositions. Prove that any expression of an even permutation in this way must use an even number of neighbor transpositions, and the same for odd permutations.

(d) Conclude that composing two odd or two even permutations gives an even permutation, and that composing one odd and one even permutation (in either order) gives an odd permutation.

2. Prove that, if $v_1, \dots, v_k$ are vectors in V, then the pure wedge $v_1 \wedge \cdots \wedge v_k \in \Lambda^k V$ is zero if and only if the vectors are linearly dependent. [Hint: We already did a lot of the work for the harder direction while computing the dimension of $\Lambda^k V$.]

3. Prove that if $\dim V = n$, then there are no nonzero alternating k-linear forms on V for $k > n$, and therefore in this case $\Lambda^k V = 0$.

4.   (a) Construct a linear map $s : V^* \otimes W^* \to (V \otimes W)^*$. Prove that it's an isomorphism when V and W are finite-dimensional. [Hint: An argument very similar to the one we used for h in the previous section will also work here.]

(b) Show in the same way as before that there is a linear map from $(V^*)^{\otimes k}$ to $(V^{\otimes k})^*$ which is an isomorphism when V is finite-dimensional.

(c) Given a k-linear form p, we can produce an alternating k-linear form using a process much like the one we used in the dimension computation in this section: define

$$\bar{p}(v_1, \dots, v_k) = \sum_{\pi} \operatorname{sgn}(\pi) p(v_{\pi(1)}, \dots, v_{\pi(k)}).$$

Prove that $\bar{p}$ is in fact alternating, and that this gives a surjective map from $(V^{\otimes k})^*$ to $(\Lambda^k V)^*$. [Hint: If p is already alternating, what is $\bar{p}$?]

(d) We can combine this surjection with the map we built in part a to get a surjection $(V^*)^{\otimes k} \to (\Lambda^k V)^*$. Prove that this descends to a surjection $\Lambda^k(V^*) \to (\Lambda^k V)^*$ which is an isomorphism when V is finite-dimensional.

(e) Conclude that if $f : V \to V$ is a linear map then $\det f = \det(f^*)$. [Hint: What is the dual of a scalar map?]

5. Suppose we have a linear map $f : V \to V$ and a basis $e_1, \dots, e_n$ of V. Pick one of these basis vectors e_i and split off the e_i coefficient of each $f(e_j)$, writing

$$f(e_j) = \alpha_j e_i + v_j,$$

where v_j is a linear combination of the basis vectors other than e_i. Use this to find a formula for the determinant of f in terms of the determinants of smaller $(n-1) \times (n-1)$ matrices formed from the entries of the matrix for f. (This formula is sometimes called **Laplace expansion**.)

6. Suppose we have a linear map $f : V \to V$ where $\dim V = n$.

   (a) Generalize the computation of $\det f$ from this section to figure out the entries of the matrix for $\Lambda^k f$ when $k < n$.

   (b) Use this to prove that a matrix T has rank less than k if and only if each $k \times k$ submatrix — that is, the matrix you get by selecting k different rows and k different columns from T — has determinant 0. (In particular, taking $k = n$, f is invertible if and only if its determinant is nonzero.)

   (c) Generalize this to linear maps $f : V \to W$, where V and W might have different dimensions.